Behavior of passenger train delays and estimation of network effects

Harrod, Steven; Cerreto, Fabrizio; Nielsen, Otto Anker; Pournaras, Georgios; Nielsen, Bo Friis

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

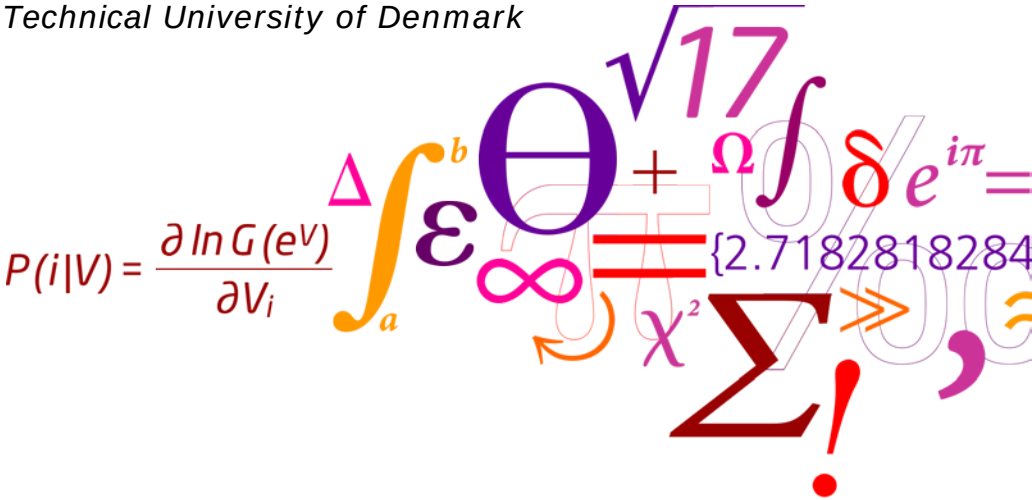
Behavior of passenger train delays and estimation of network effects

5th International Symposium on Railway Operations Research, 2018

Beijing Jiaotong University

Steven Harrod[a], Fabrizio Cerreto[d], Otto Anker Nielsen[a], Georgios Pournaras[c], Bo Friis Nielsen[b]

[a] *Department of Management Engineering, Technical University of Denmark*
[b] *Institut for Matematik og Computer Science, Technical University of Denmark*
[c] *Ansaldo STS - a Hitachi company group*
[d] *Banedanmark (Rail Net Denmark)*



Innovation Fund Denmark
RESEARCH, TECHNOLOGY & GROWTH

DTU Management Engineering

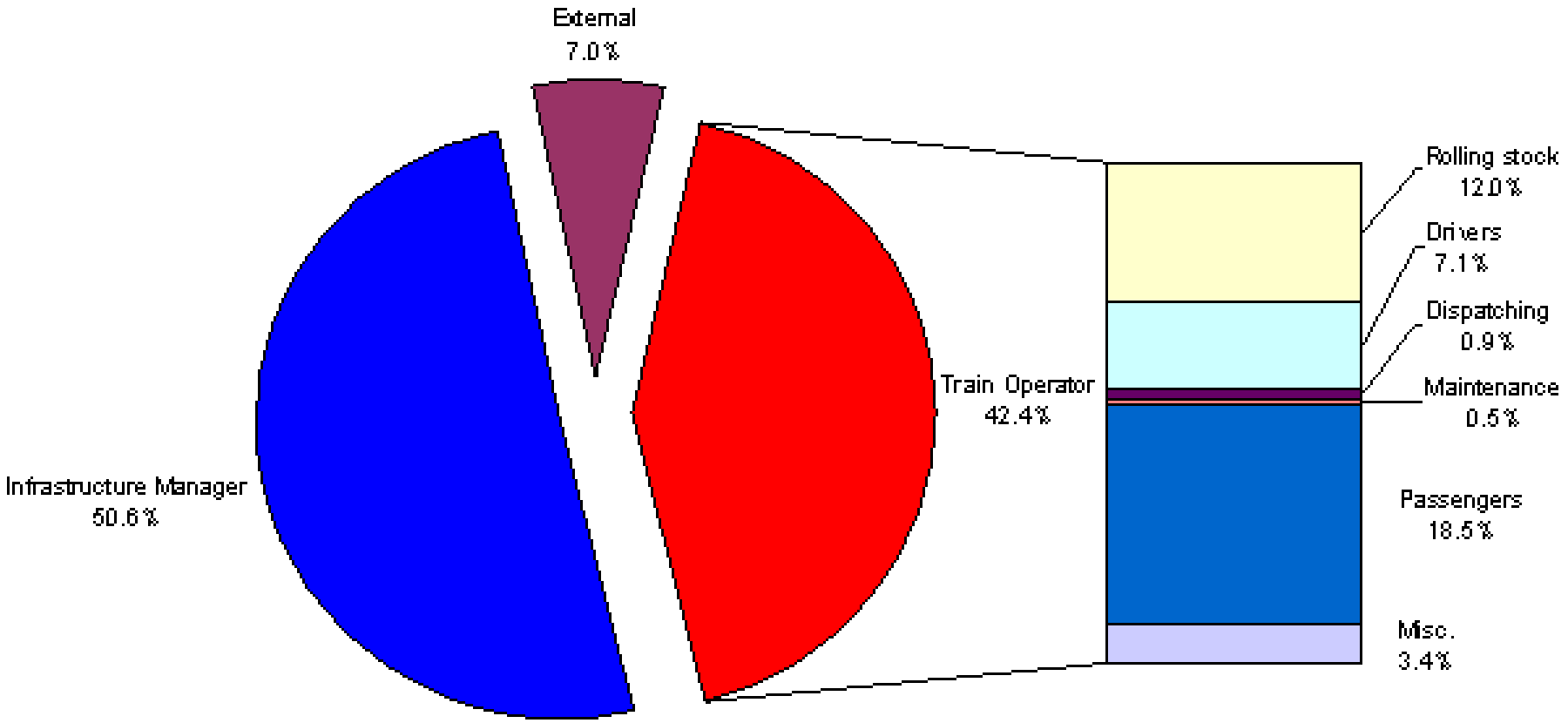
Today's Program

- Highlights From Punctuality Research at DTU
- Background: Management Has Value
- The Data May Not Be What It Seems
- Timetables Are a Limited Tool
- Look For the Thread

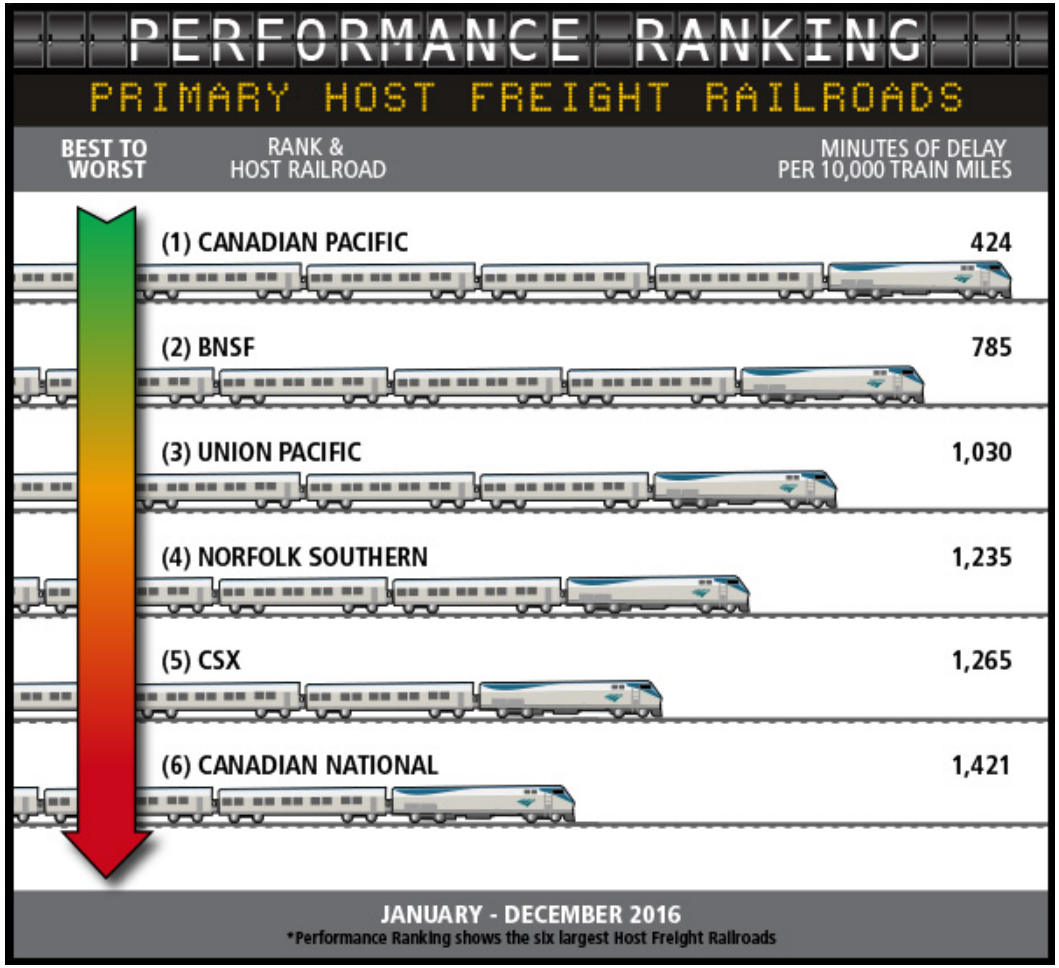
Part of IPTOP: Integrated Public Transport Optimization and Planning

Funded by Innovation Fund Denmark

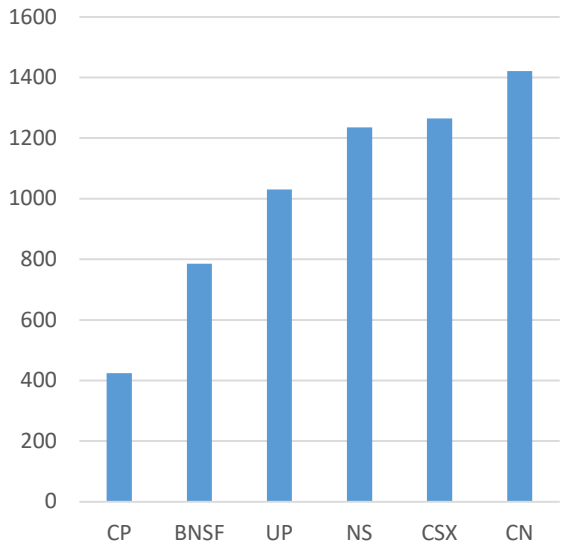
Train Delays, Typical



Management Is a Significant Factor

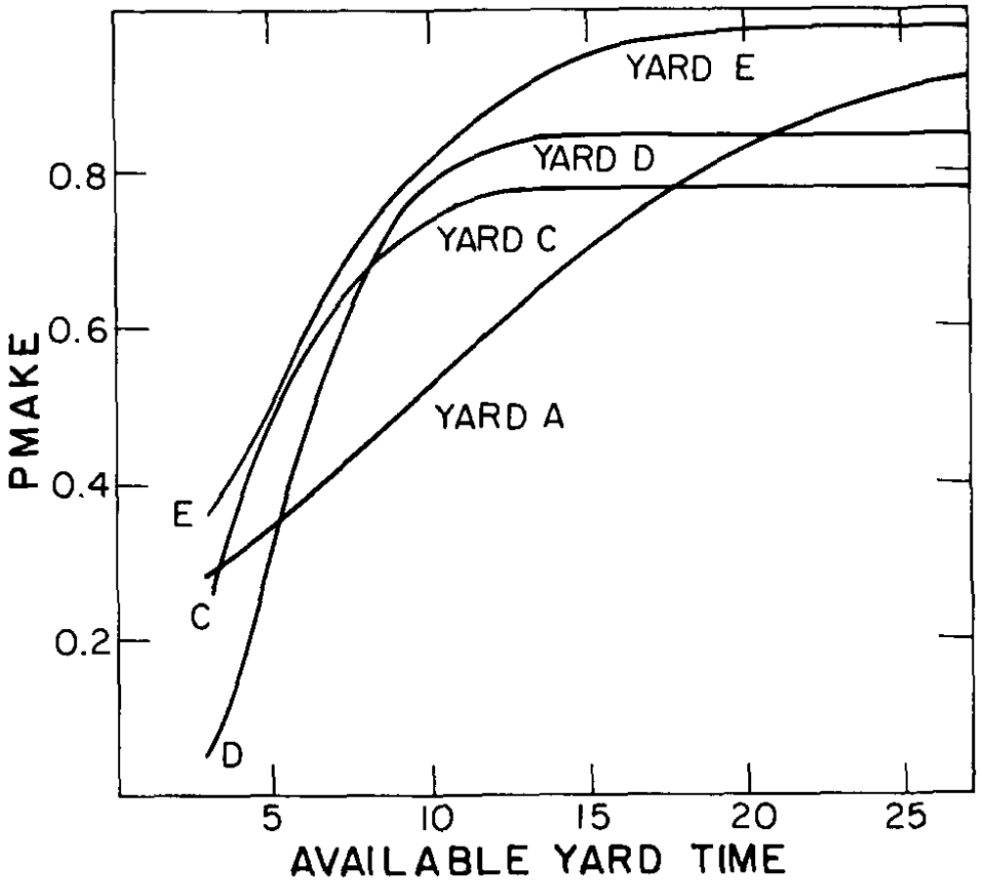


Minutes of delay per 10.000 train miles



Amtrak passenger train delay 2016

The Pmake Wagonload Freight Connection Probability



Martland, C.D. (1982) Pmake Analysis: Predicting Rail Yard Time Distributions Using Probabilistic Train Connection Standards, Transportation Science v.16 (4), pp. 476-506

Topic 1: Fitting Distributions to Data Who Cares?

- Primary delays – for simulation and modeling
- Aggregate delays
  - –for performance estimation and forecasting
  - –for validation of simulation models
- Both of management interest

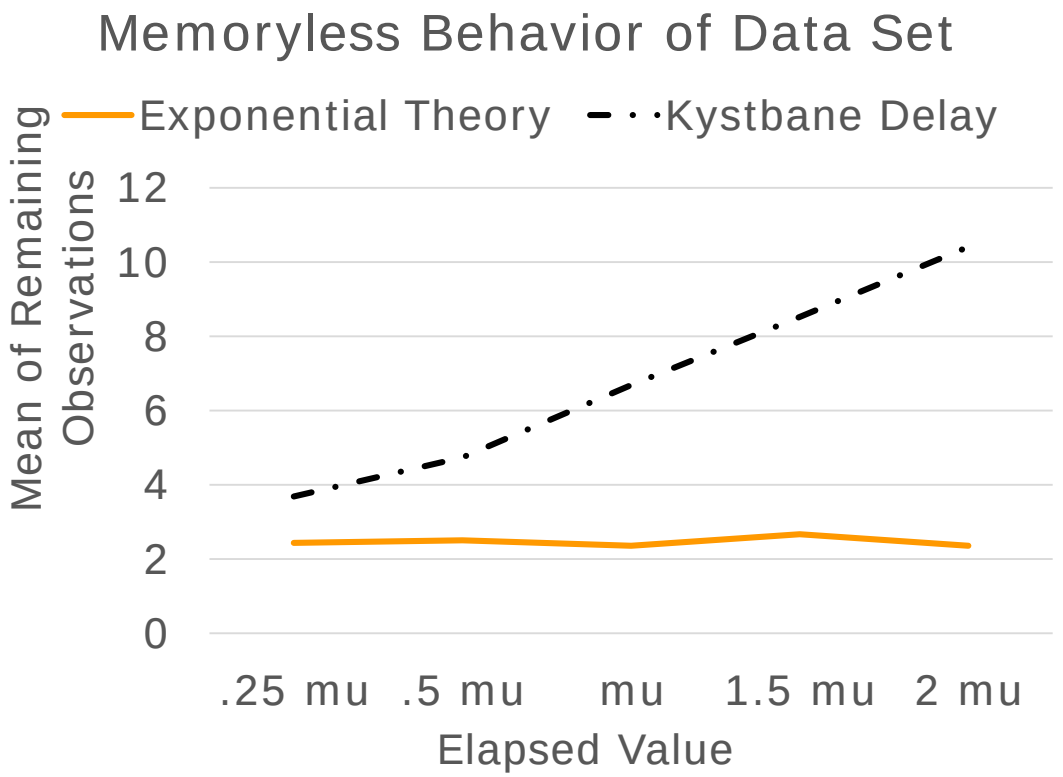
Brown and Cudeck 1992

- "Statistical goodness-of-fit tests are often more a reflection on the size of the sample than on the adequacy of the model"

Sample sizes of less than 200 will frequently satisfy statistical fit tests when in fact the fit is incorrect. Today's larger data sets are giving different results and changing our understanding of fundamental behavior.

Second Problem Process is Not Memoryless

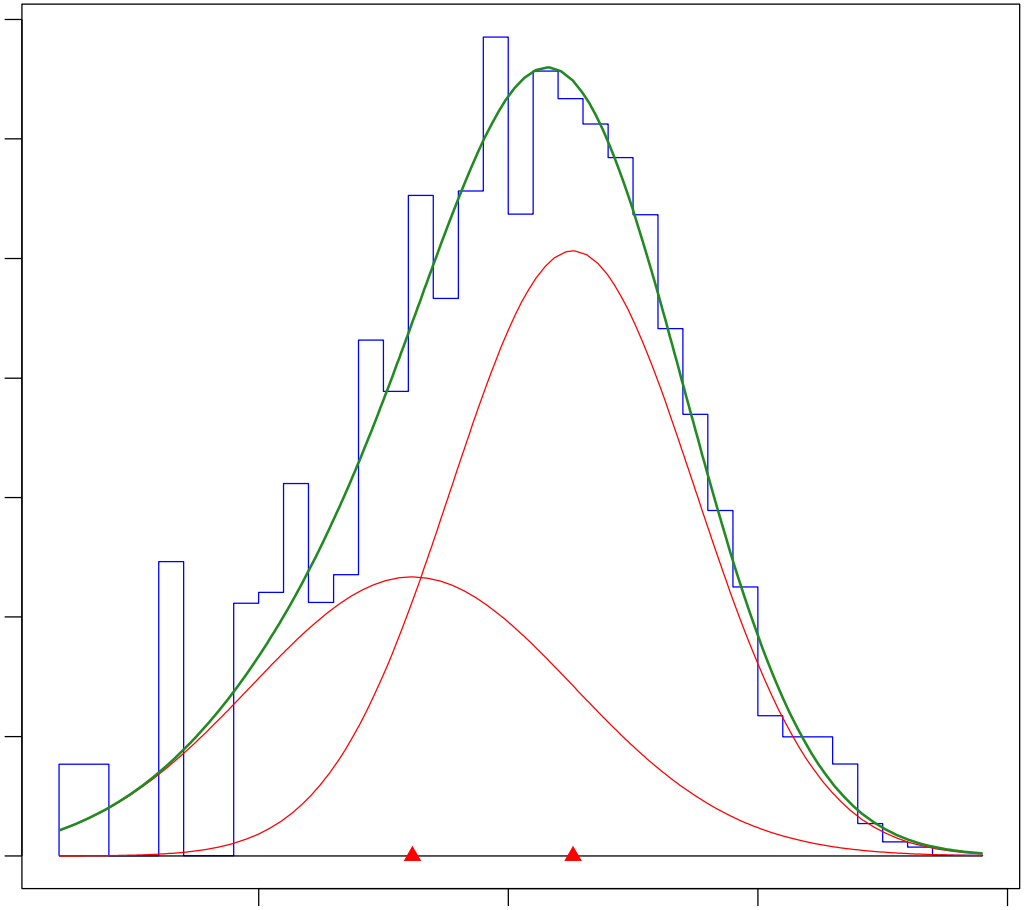
- An exponential process should be memoryless
- Delay data is not



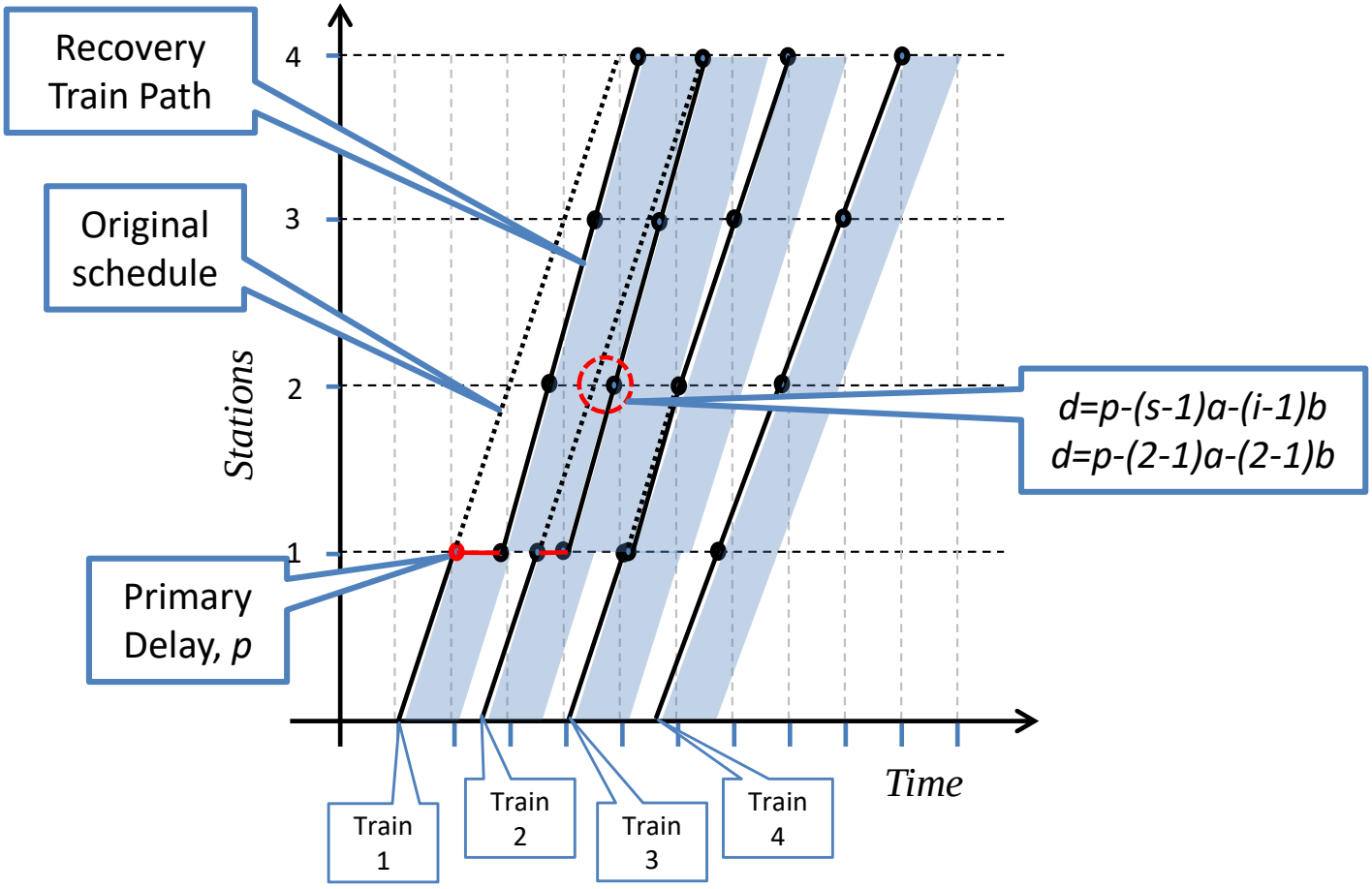
Mixed Distribution Fit

25 000 data points from Danish Kystbane

Best fit: transform to log base 10, mixed normal distribution

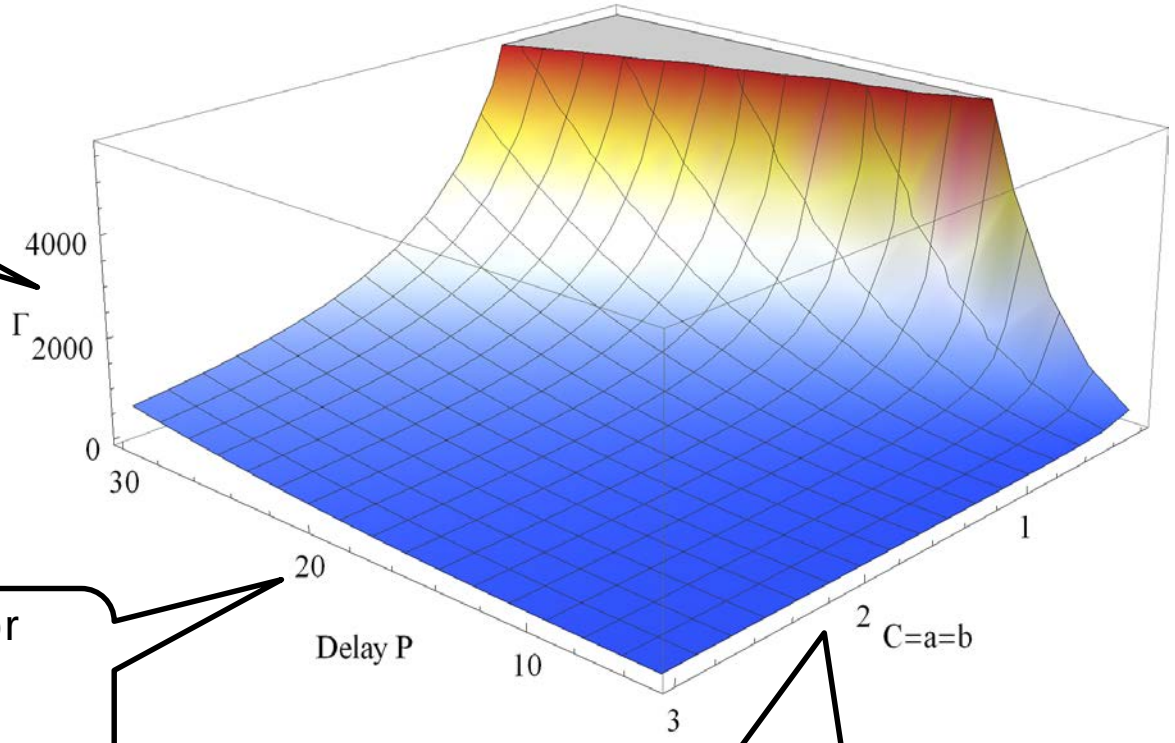


Topic 2:
Modeling Network Delay



The Network Aggregate Effect of a Single Delay is Polynomial

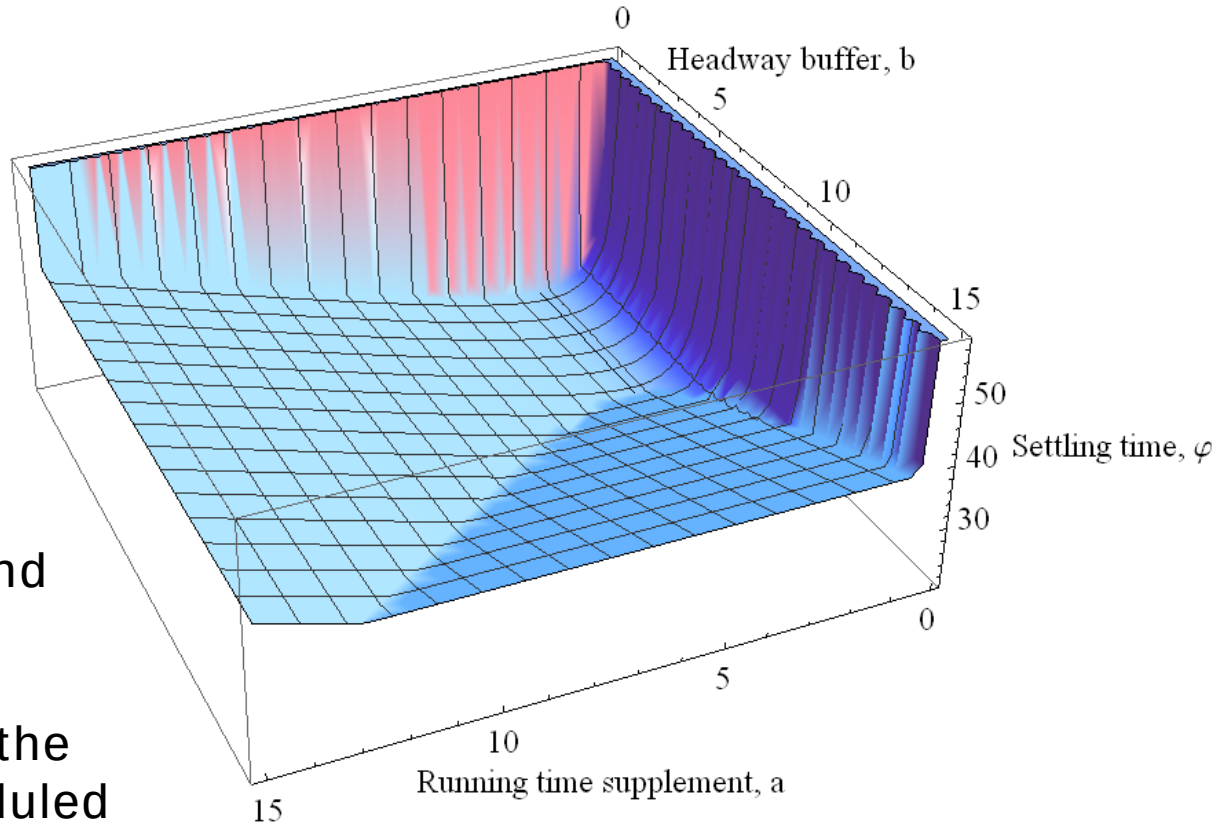
Total minutes of delay over all stations all trains



Delay minutes for one train at one point

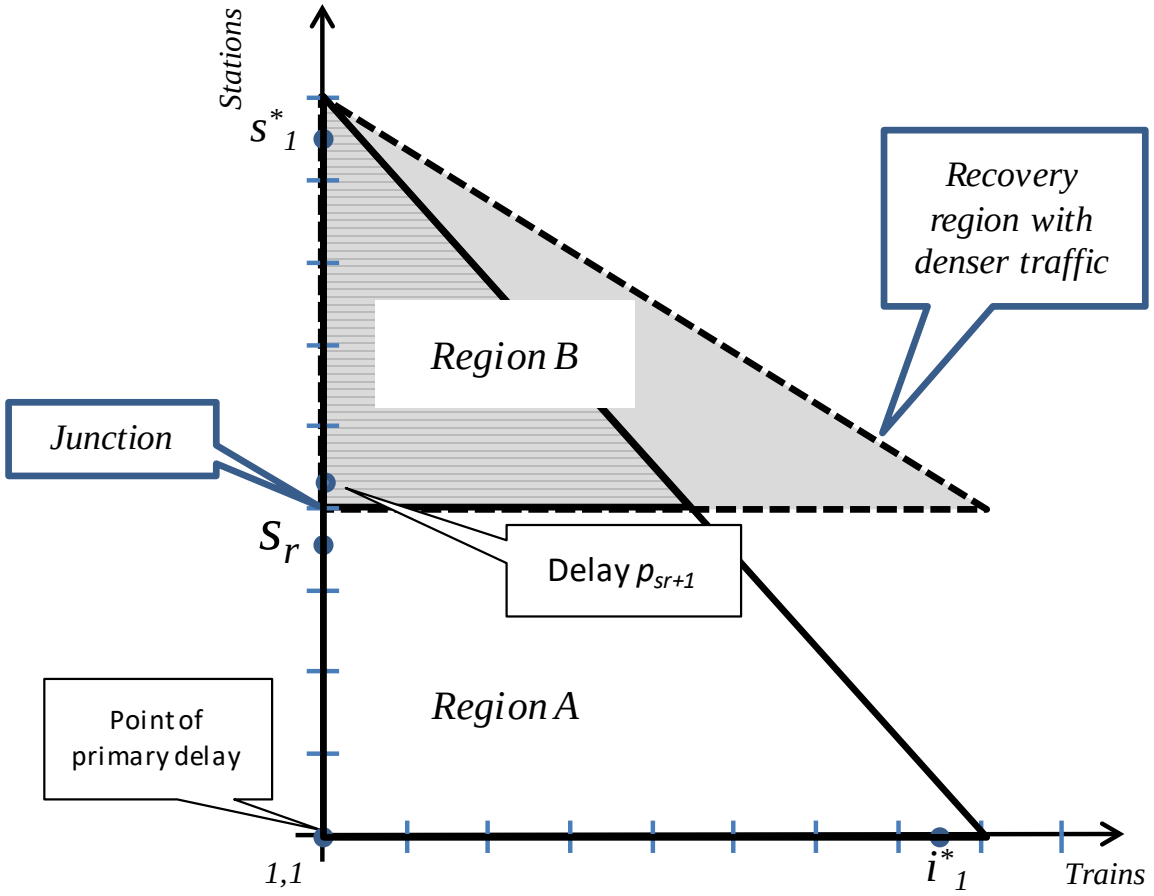
Size of timetable supplement and buffer, both equal

Settling Time Can Actually Be Worse with Too Much Slack



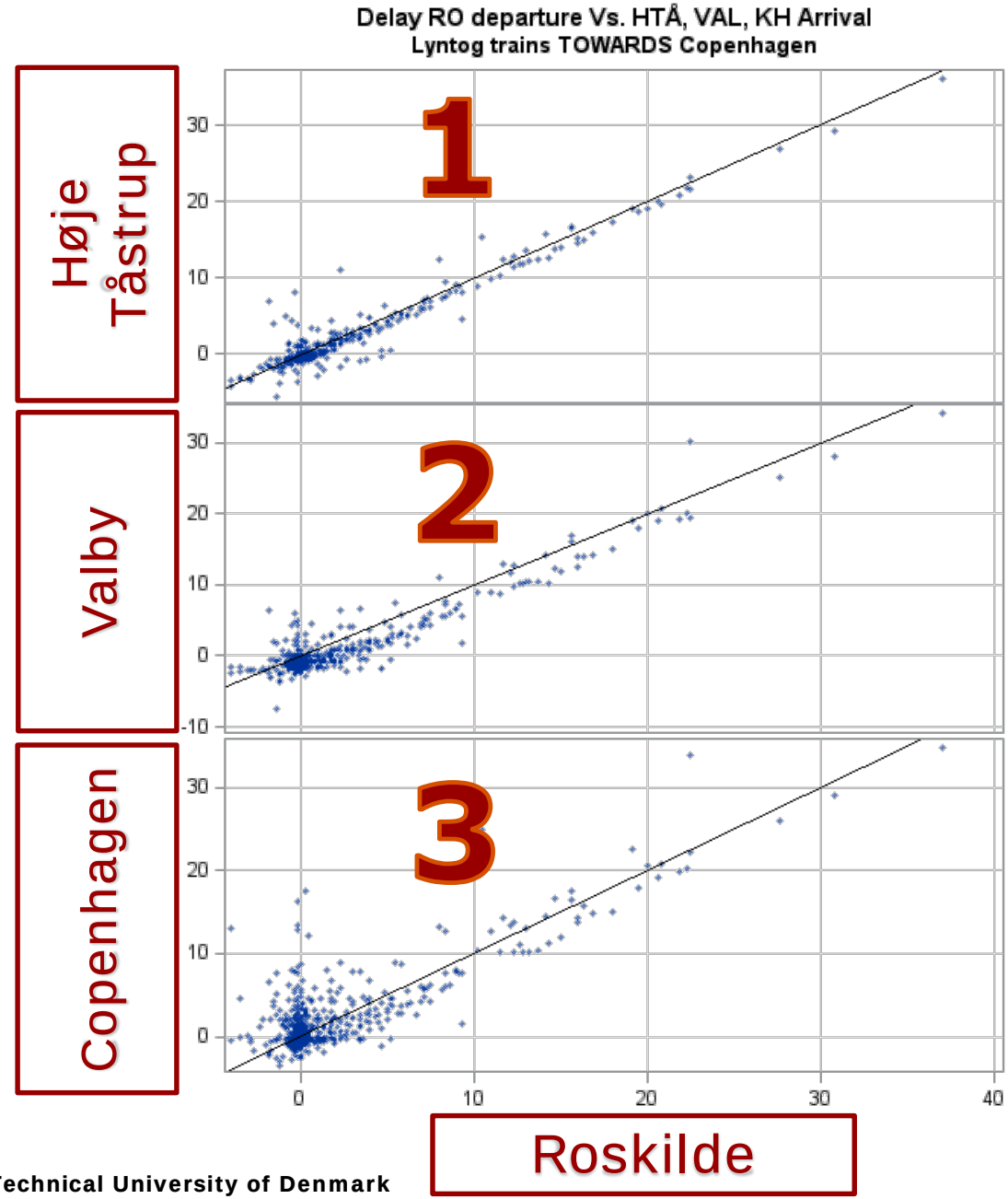
Larger supplements and buffers not only make journeys longer, they interfere with getting the network back to scheduled operation

Aggregate Delay Network Effects Can Be Decomposed



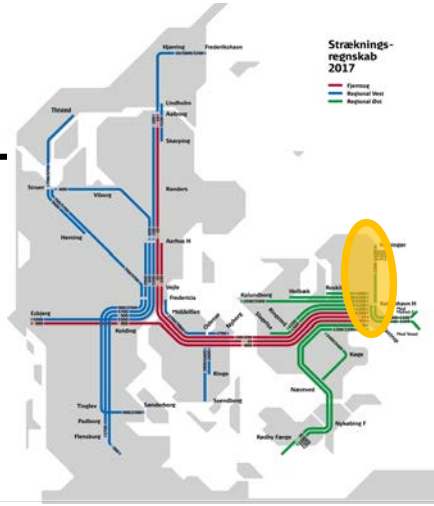
Topic 3: And Yet, Real Data Confounds Expectations

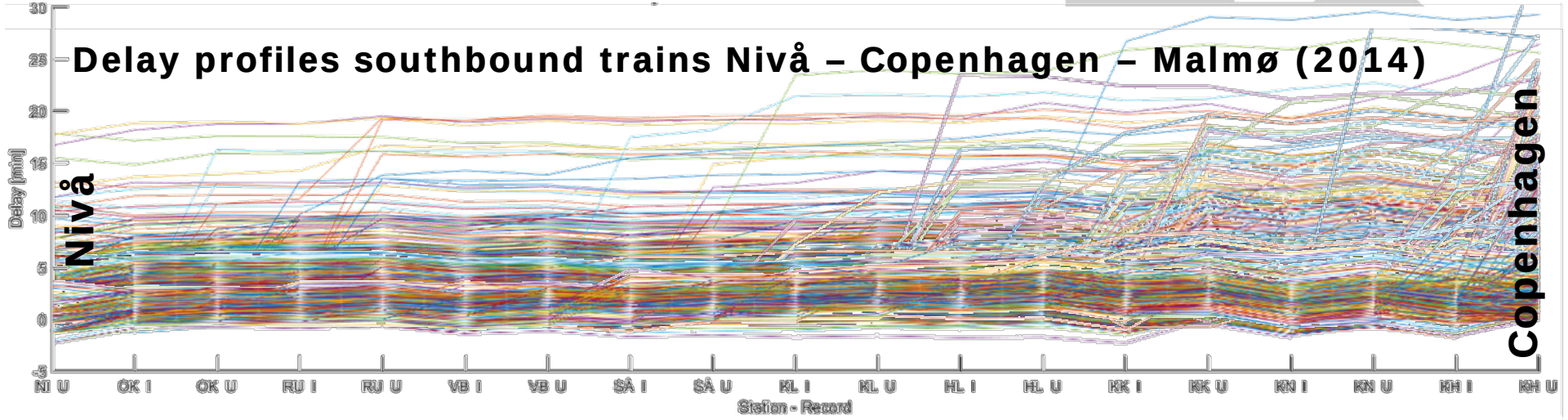
Trains mysteriously become late at end of journey



Deeper Causal Analysis is Needed

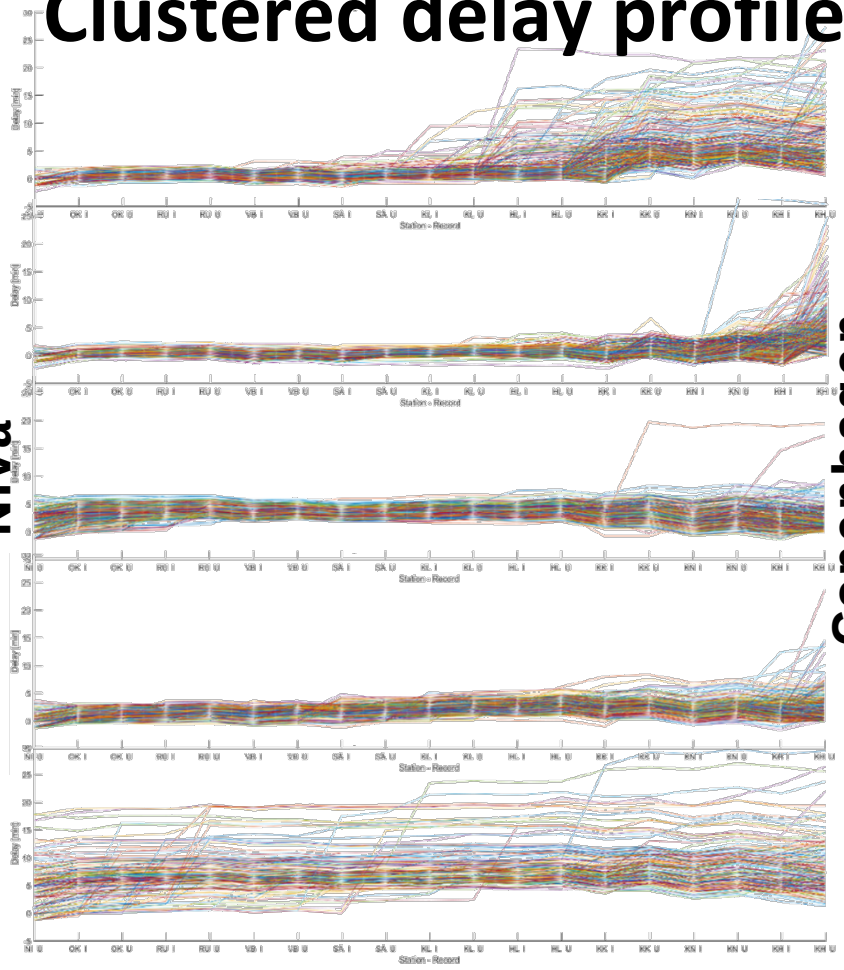
Kystbanen Helsingør-Copenhagen





16

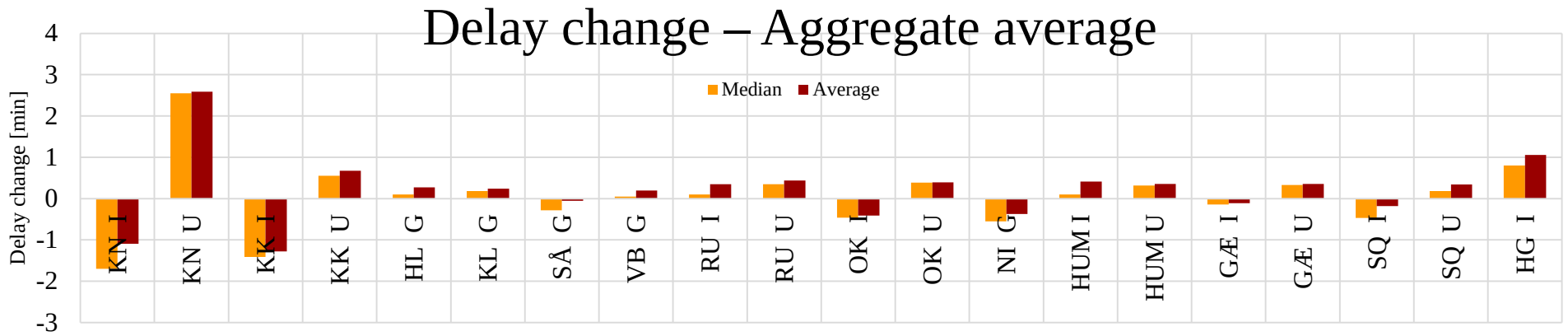
Clustered delay profiles

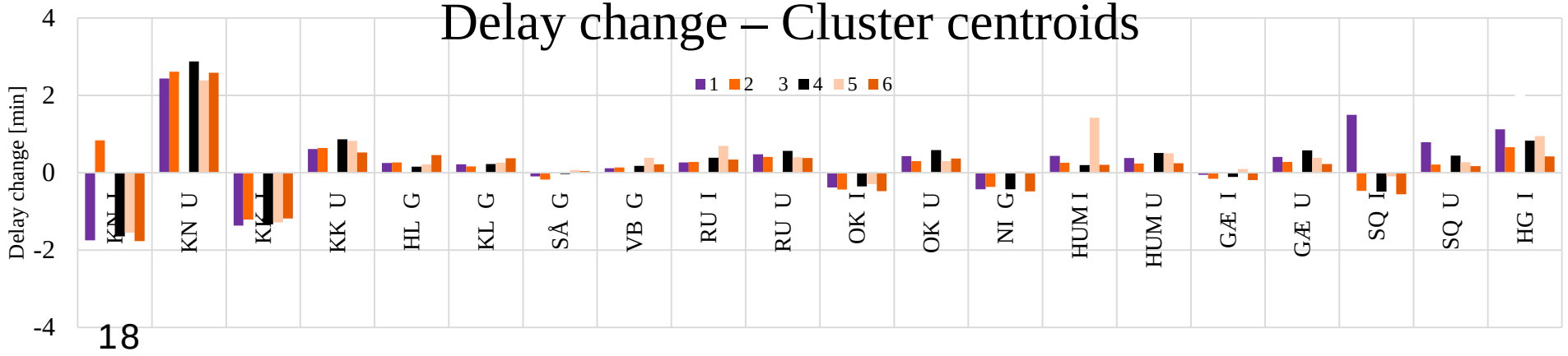


- Delays from KL to Copenhagen
- Delays leaving Copenhagen
- Punctual (<5 min) constant
- Small delays, increasing towards Copenhagen
- Larger delays

17

Delay change profiles





18

Summary

- We can predict deterministic performance of train delays
- Collected data does not behave as predicted
- More focus needs to be on structural and managerial causes of delay

Thank you!

Tak!